

SOSNet:一种非对称编码器-解码器结构的 非小细胞肺癌CT图像分割模型

谢娟英,张凯云

(陕西师范大学计算机科学学院,陕西西安 710119)

摘要: 非小细胞肺癌严重损害人类健康,早期非小细胞肺癌CT(Computed Tomography)图像中的肿瘤结节体积小,不易发现,极易造成漏诊和误诊.为了精确分割非小细胞肺癌CT图像中的小体积肿瘤结节,本文提出SOSNet(Small Object Segmentation Networks)自动分割模型,利用ResNet(Residual Network)基础层和空洞卷积构造非对称编码器-解码器结构作为分割主网络,利用轴向取反注意力模块ARA(Axial Reverse Attention)逐步擦除背景中对分割有影响的区域,再使用结构细化模块SR(Structure Refinement)对主网络输出的粗略特征图进行结构细化,从而实现非小细胞肺癌肿瘤结节分割.在非小细胞肺癌公开数据集的实验测试表明,本文提出的小目标自动分割模型SOSNet可以有效分割出非小细胞肺癌CT图像中的小体积肿瘤结节,其mDice(mean-Dice)、mIoU(mean Intersection over Union)、Sensitivity、F1、Specificity、平均绝对误差MAE(Mean Absolute Error)均优于当前最先进的小目标分割模型CaraNet(Context Axial Reverse Attention Network).

关键词: 小目标分割;非小细胞肺癌;非对称编码器-解码器;结构细化;轴向取反注意力;CT图像;深度学习;卷积

基金项目: 国家自然科学基金(No.62076159, No.12031010, No.61673251);中央高校基本科研业务费项目(No.GK202105003)

中图分类号: TP181;TP391

文献标识码: A

文章编号: 0372-2112(2024)03-0824-14

电子学报URL: <http://www.ejournal.org.cn>

DOI:10.12263/DZXB.20220853

SOSNet: An Asymmetric Encoder-Decoder Structure Model for Automatic Segmenting Non-Small Cell Lung Cancer CT Images

XIE Juan-ying, ZHANG Kai-yun

(School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi 710119, China)

Abstract: Non-small cell lung cancer (NSCLC) will imperil human health seriously. The tumor nodules at the early stage of NSCLC are so small that it is very difficult to detect them in the CT (Computed Tomography) images, which will easily lead to the missed diagnosis and misdiagnosis of NSCLC. To automatically segment the small tumor nodules in CT images of NSCLC accurately, the SOSNet (Small Object Segmentation Networks) model is proposed. The ResNet (Residual Network) base layer and the dilated convolution are adopted to construct the asymmetric encoder-decoder structure to be the segmentation main network of SOSNet. The ARA (Axial Reverse Attention) module is adopted to gradually erase those structures which may influence the segmentation results from the background. Then the SR (Structure Refinement) module is used to refine the rough feature maps outputted by the main network, so as to achieve the segmentation for NSCLC tumor nodules. Experimental results on the open access NSCLC datasets demonstrate that the proposed SOSNet model can effectively segment small volume tumor nodules in CT images of NSCLC. It is superior to the state-of-the-art small object segmentation model of CaraNet in terms of mDice (mean Dice), mIoU (mean Intersection over Union), Sensitivity, F1, Specificity and MAE (Mean Absolute Error), respectively.

Key words: small object segmentation; non-small cell lung cancer; asymmetric encoder-decoder framework; structure refinement; axial reverse attention; CT images; deep learning; convolution

Foundation Item(s): National Natural Science Foundation of China (No.62076159, No.12031010, No.61673251); Fundamental Research Funds for the Central Universities (No.GK202105003)

1 引言

肺癌被称为癌症之王,是世界上最常见的恶性肿瘤之一,其发病和死亡率最高,并呈逐渐上升趋势^[1].据中国国家癌症中心数据显示,非小细胞肺癌 NSCLC (Non-Small Cell Lung Cancer) 占有所有肺癌的 80%,由于早期肺癌并没有明显的临床症状,大约有 75% 的非小细胞肺癌患者在发现时已处于中晚期,因此,越早发现越早治疗对于挽救病人生命具有非常重要的意义.在肺癌筛查和治疗阶段,胸部计算机断层扫描(Computer Tomography, CT)十分重要,有经验的医生可以从 CT 中发现病变区域的位置和范围,并进行诊断和治疗.但是,诊断结果依赖于医生的专业知识和经验,费时费力.另外,胸部 CT 中包含了气管、血管、胸骨以及其他众多干扰诊断分析的实质,给诊断带来一定影响和困难.而且,CT 中病变区域以结节为主,结节体积小,很容易被忽略.因此,借助人工智能技术精确分割肺部结节和肿瘤对非小细胞肺癌筛查和治疗具有重要作用.

随着 Long 等人^[2]2015 年提出全卷积神经网络 FCN (Fully Convolution Networks),语义分割取得了巨大进步.针对医学图像语义分割,Ronneberger 等人^[3]在 FCN 基础上提出左右对称的编码器-解码器结构 U-Net,利用对称的编、解码结构及跳跃连接对细胞图像进行分割,奠定了编码器-解码器结构在医学图像分割中的重要地位.基于 U-Net 模型的肺部 CT 图像分割得到关注,如谢等人^[4]通过使用 XR(X ResNet)模块和 MSF(Multi-Scale Features fusion module)模块对 U-Net 结构进行改进,实现了对新冠肺炎肺部 CT 图像感染区域的有效分割.Tong 等人^[5]将残差结构加入 U-Net 模型,实现了肺结节的精确分割;Rocha 等人^[6]为节省计算成本并提高网络存储效率,在 U-Net 模型引入池化索引操作^[7],并加入到最大池化和上采样之间,使用反池化技术进行上采样,但其结节分割精度没有得到明显提升.Singadkar 等人^[8]提出一种端到端分割不同大小肺结节的反卷积残差网络,利用长跳连接将卷积和反卷积进行部分连接,捕获全局分辨率并补充池化操作丢失的空间信息.Wang 等人^[9]将原始 512×512 大小的 CT 图像划分为 64×64 大小的 64 个小块图像,然后从中挑选出有结节的图像块送入分割模型中进行分割,虽然该方法取得了不错的分割效果,但是会耗费大量计算资源.Ait 等人^[10]通过在 U-Net 的输入和输出层加入密集网络并利用空洞卷积扩大感受野,生成多尺度特征来改进 U-Net,以发现肺癌小型结节,这一网络结构对小型肺结节的分割准确率提高了 20%.

以上基于卷积神经网络的肺部图像分割,本质上是基于图像像素的端到端分类,直接从肺部 CT 图像中分割出结节.肺结节分割还可分解为多步进行,首先从

肺部 CT 图像分割出双肺,然后再进一步分割出结节,如 Blanc 等人^[11]提出三阶段肺部 CT 图像处理,先使用 3D U-Net 对 CT 图像中的肺区进行分割,然后使用 3D Retina-U-Net 检测肺区中的结节,最后用支持向量机对检测到的结节进行分类.

尽管非小细胞肺癌 CT 图像分割已得到诸多学者关注,但非小细胞肺癌 CT 图像中体积过小的肿瘤结节分割存在挑战.首先,肺结节直径大小一般在 1.5~15 mm 之间,在 CT 图像的体积占比非常小,设计一种对大、小结节分割效果都好的模型很困难;其次,较小体积的肿瘤结节需要更加精细化的边界分割,才能保证分割结果的精确性;然而,大多数深度神经网络忽略小目标对模型分割性能的影响,只关注模型的整体分割性能.

因此,针对非小细胞肺癌 CT 图像中体积较小的肿瘤结节分割,本文提出小目标分割模型 SOSNet(Small Object Segmentation Networks).采用 ResNet(Residual Network)基础层作为 SOSNet 模型的编码器进行特征提取,采用非对称空洞卷积解码器进行上采样,还原图像.其次,利用轴向取反注意力模块 ARA(Axial Reverse Attention)改善模型对小目标分割时的漏检和误检.最后,使用边界细化模块 SR(Structure Refinement)完善小目标的分割边界,实现更加精细化的边界分割效果.

2 相关基础

2.1 编码器-解码器结构

编码器-解码器(encoder-decoder)结构是一类深度学习框架,显著特征是端到端(end-to-end)学习.编码器-解码器结构最初被应用于机器翻译^[12],但是随着 SegNet^[7]的提出,编码器-解码器结构开始在语义分割领域流行.2015 年 Ronneberger 等人^[3]提出具有编码器-解码器结构的 U-Net 网络完成了细胞图像的端到端分割,开启了编码器-解码器结构在医学图像分割领域的应用.

语义分割领域的编码器-解码器结构如图 1 所示.编码器和解码器两部分可以完全对称,也可以不完全对称.一般来说,编码器由卷积和下采样构成,连续卷积和下采样使得特征图的尺寸不断缩小,提取到越来越丰富的低级特征和高级特征,得到具有丰富空间信息和全局信息的特征图.简单来说,编码器的主要作用是将输入图像转换成具有更低维度表征的特征图.解码器的作用与编码器相反,利用反卷积等上采样操作对编码器得到的特征图进行操作,逐步恢复特征图尺寸和细节信息,并与对应编码器提取的特征进行融合,在尽可能减少信息损失的前提下,输出与输入图像同

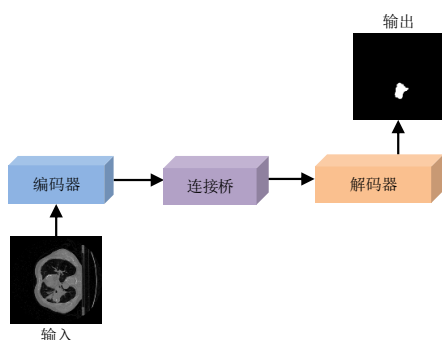


图1 语义分割中的编码器-解码器结构

尺寸的分割预测图,完成端到端的分割。

2.2 空洞卷积

空洞卷积(dilated/atrous convolution)的雏形是Holschneider等人^[13]1989年为了计算非抽样离散小波变换,提出的“带孔算法”,但空洞卷积^[14]这个术语到2015年才首次出现。Yu等人^[15]将带上采样或放大功能的滤波器操作称为扩张卷积,但没有将其用于深度学习模型。深度学习中,为了获得更大的感受野,需要扩大卷积核尺寸,随着卷积核尺寸的扩大,训练参数量会急剧增加,容易造成过拟合,为了避免过多参数和过拟合问题,空洞卷积被引入。DeepLab系列网络^[16-19]将空洞卷积很好地融入到深度学习模型,大大优化了端到端分割网络的语义分割性能。另外,空洞卷积技术还被广泛用于目标检测^[20,21]、视觉问答^[22]等任务,并取得了良好性能。

空洞卷积的主要作用是扩大感受野,以获得更多上下文信息,与普通卷积不同的是空洞卷积具有空洞率,当空洞率等于1时,空洞卷积就是普通卷积,当空洞率大于1时,卷积核会对特征图以空洞率减1进行间隔采样,这样就会获得更大的感受野,空洞卷积的卷积核大小计算如式(1),对应空洞卷积感受野大小的计算公式如式(2)所示。

$$K_A = K_o + (K_o - 1) \times (\text{rate} - 1) \quad (1)$$

$$RF_{i+1} = RF_i + (K_A - 1) \times S_i \quad (2)$$

其中, K_A 和 K_o 分别表示空洞卷积核的大小和对应普通卷积核的大小, rate 表示空洞率, RF_{i+1} 和 RF_i 分别表示当前层和上一层的感受野大小, $S_i = \prod_{j=1}^i \text{stride}_j$ 表示当前层以前所有层的步长之积, stride_j 表示第 j 层的步长。图2展示了卷积核大小为 3×3 ,各层步长为1时,从左到右不同空洞率的空洞卷积进行连续卷积对应的感受野。

2.3 自注意力机制

自注意力机制(self-attention mechanism)是一种注意力机制,由Google机器翻译团队的Vaswani等人^[23]于2017年提出,用来学习文本表示。此后,自注意力机制

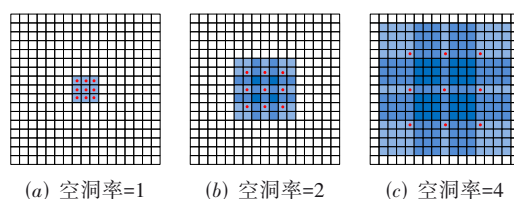


图2 步长为1时不同空洞率的空洞卷积连续卷积的感受野

成为研究热点。在计算机视觉领域,自注意力机制主要是为增强特征之间的相关性而设计。为了捕获更加丰富的全局信息,Wang等人^[24]提出基于图像像素间协方差的自注意力机制,通过关注特征图的所有像素,并在嵌入空间取其加权平均来表示特征图某位置处的响应,解决长距离依赖,建立所有像素之间的联系,从而捕捉特征图的全局信息。Zhang等人^[25]将自注意力机制引入生成对抗网络(Generative Adversarial Network, GAN)^[26],解决普通卷积的长距离依赖问题,实现在减少参数量的同时,扩大卷积核的感受野。在语义分割领域,Fu等人^[27]提出基于自注意力机制捕获上下文依赖关系,解决场景分割问题的DANet(Dual Attention Network)网络,利用位置注意力机制和通道注意力机制双重注意机制增强全局特征融合与各位位置特征的联系。

非小细胞肺癌CT图像中的肺结节目标较小,需要结合全局信息才能更好地进行分割,自注意力机制可以有效增加语义特征之间的相关性,增强全局特征融合,对于精确分割目标提高模型整体分割性能十分有效。

3 SOSNet模型

本节首先介绍提出的SOSNet模型的整体结构,然后分别详细阐述编码器和解码器模块,接着介绍用于获得更加精确分割边界的结构细化模块SR,并介绍轴向取反注意力模块ARA,以及训练模型使用的损失函数。

3.1 SOSNet模型结构概述

本文提出的SOSNet模型由非对称编码器-解码器结构、轴向取反注意力模块ARA、结构细化模块SR三部分组成,其网络结构如图3所示。非对称的编码器-解码器结构构成从输入图像获得粗略分割特征图的主网络;轴向取反注意力模块ARA用来增加模型对于小目标的分割精度;结构细化模块SR用来细化主网络获得的粗略分割特征图。

非对称编码器-解码器结构组成的主网络中,编码器由不同数量的ResNet^[28]层、批归一化BN(Batch Normalization)以及ReLU(Rectified Linear Unit)组成,解码器则由 3×3 大小,空洞率为2的空洞卷积、批归一化BN以及ReLU激活函数组成,其中编码器部分采用 2×2 大

小的最大池化进行下采样,解码器部分采用双线性插值进行上采样,在编码器和解码器之间采用了3组由 3×3 大小,空洞率为2的空洞卷积和批归一化BN以及ReLU激活函数组成的模块作为连接桥.在对应的编码器模块和解码器模块之间加入了轴向取反注意力模块ARA.主网络输出的分割结果(粗略特征图)作为输入进入结构细化模块SR进行边界结构细化,最终模型的输出为结构细化模块SR的输出.

3.2 非对称编码器-解码器主干网

编码器-解码器结构能够同时获得图像中高层、全局、上下文以及底层细节信息,所以本文使用该结构作

为特征提取的主要架构.编码器部分与BASNet (Boundary-Aware Segmentation Network)模型的主分割框架^[29]相同,主要由一个 3×3 大小的输入卷积层和六个由ResNet层构成的特征提取块组成:其中输入卷积层和前四个特征提取块均来自ResNet-34^[28],与ResNet-34不同的是将 7×7 大小,步长为2的卷积改为 3×3 大小,步长为1的卷积,这样的改进能够使网络获得更高分辨率的特征图.在第四个特征提取块之后,又增加了两个分别由三个基本ResNet层组成的特征提取块,分别在第二个到第六个特征提取块的最后一层ResNet层后加入批归一化BN和ReLU激活函数,并使用 2×2 最大池化(MaxPool)进行下采样.

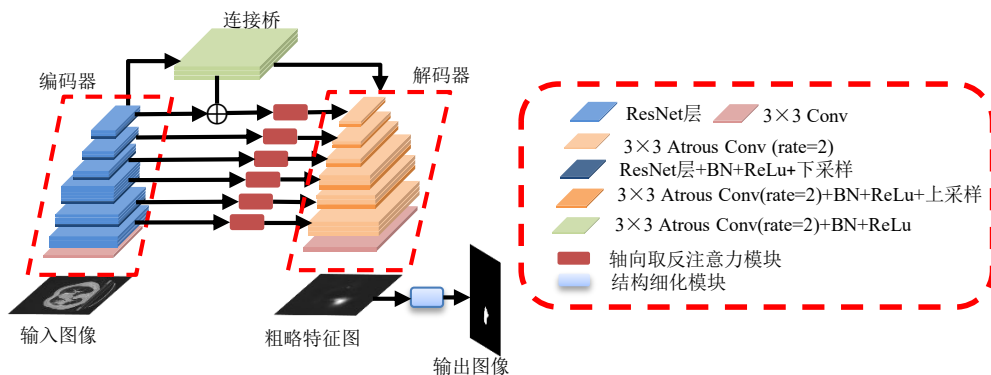


图3 SOSNet网络结构示意图

编码器和解码器之间还有一个三层的连接桥,由大小 3×3 、空洞率为2的空洞卷积组成,目的是获得更加丰富的全局信息,在每一层卷积之后都有一个批归一化BN处理和ReLU激活函数.

解码器部分由一个 3×3 大小的输出卷积层和六个由空洞卷积构成的信息恢复块组成.其中,使用空洞卷积可提升模型对广域信息的建模能力,避免模型过度聚焦局部特征,导致目标区域或边界分割不充分.为避免空洞卷积在解码过程中出现严重的“网格效应”,解码器部分的空洞卷积均采用空洞率为2的 3×3 大小的小卷积核.解码器的前三个信息恢复块均由两层空洞卷积构成,目的是为了减少空洞卷积操作在特征图尺寸较小情况下引起的信息丢失,同时可保证解码器在解码初期兼顾局部和全局视野,第四到第六个恢复块均由三层空洞卷积构成.为了逐步恢复特征图尺寸,在解码器的第一个恢复块到第五个恢复块的最后一层空洞卷积后加入批归一化BN和ReLU激活函数,并使用双线性差值进行上采样.

3.3 结构细化模块SR

为了使模型能获得小目标更准确的结构分割效果,结合编码器-解码器结构可以获得底层细节信息的特点,提出基于编码器-解码器结构的结构细化模块

SR,如图4所示.

图4所示的结构细化模块与本文主网络类似,编码器部分使用一层 3×3 大小的普通卷积做输入层,使用四组 3×3 大小的普通卷积、批归一化BN和ReLU激活函数做特征提取,并使用最大池化做下采样;解码器部分也使用四组 3×3 大小空洞率为2的空洞卷积、批归一化BN和ReLU激活函数并使用双线性插值做上采样恢复特征图,最后使用一层 3×3 大小的普通卷积做输出层.同时,在输出层之后使用跳层连接将输入的粗略特征图与输出层的特征图进行相加,得到结构细化特征图,也就是整个网络输出的分割结果.

结构细化模块SR类似残差结构^[28],其原理是逐步细化粗略预测图像来改善分割结果,所以结构细化模块可以定义为式(3).

$$\mathbf{F}_{\text{refined}} = \mathbf{F}_{\text{coarse}} + \mathbf{F}_{\text{residual}} \quad (3)$$

其中, $\mathbf{F}_{\text{refined}}$ 表示结构细化特征图, $\mathbf{F}_{\text{coarse}}$ 表示粗略特征图, $\mathbf{F}_{\text{residual}}$ 表示粗略特征图经过结构细化模块学习,逐步细化后得到的细化特征图.

3.4 轴向取反注意力模块ARA

轴向取反注意力模块ARA如图5所示,主要包含两个分支,第一个分支为图像取反路径,其输入 T 为图像的真实分割图(Ground Truth),主要作用是逐步擦除

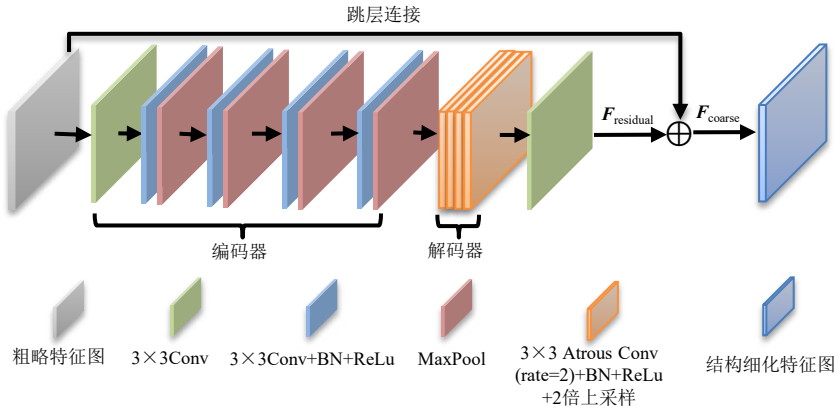


图4 结构细化模块SR

图像背景,去除干扰分割的特征信息;另一个分支为轴向注意路径,其输入为主分割网络的相应特征图 F , 主要作用为加速计算和获得全局注意力. 轴向取反注意力模块可以表示为式(4).

$$F_{\text{ARAM}} = F_{\text{reverse}} \otimes F_{\text{AM}} \oplus F_{\text{input}} \quad (4)$$

其中, F_{ARAM} 表示轴向取反注意力模块的输出, F_{reverse} 表示图像取反路径的输出, F_{AM} 表示轴向注意路径的输出, F_{input} 表示轴向注意路径的输入特征图 F , \otimes 表示对应元素相乘, \oplus 表示对应元素相加.

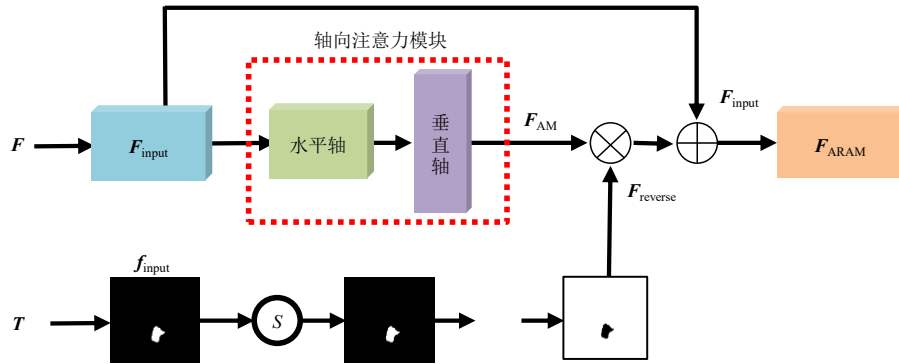


图5 轴向取反注意力模块ARA

非小细胞肺癌肺部CT图像中的肿瘤结节体积较小,而肺部CT中的血管、气管等组织很容易影响分割效果,如果将这些背景去除,只留下分割目标可以有效提升分割效果. 因此,使用了反向注意力^[30]使图像取反方式,逐步擦除图像中的背景信息,利用只含有目标的取反图像,增加模型对于小目标的分割效果. 图像取反路径可以表示为式(5).

$$F_{\text{reverse}} = 1 - S(f_{\text{input}}) \quad (5)$$

其中, F_{reverse} 表示取反得到的特征图, f_{input} 表示输入特征图, S 为 Sigmoid 函数.

轴向注意路径使用轴向注意力^[31]实现,轴向注意力是自注意力^[23]的改进,虽然自注意力可以作为卷积的补充,使网络捕获远程关系,但单独的自注意力局限于局部区域,需经过堆叠才能实现全局注意力. 而轴向注意力通过将一个二维注意力分解为水平轴和垂直轴的两个一维注意力,在扩大模型注意

力范围的同时,降低计算量. 轴向注意路径可表示为式(6).

$$F_{\text{AM}} = F_y(F_x(F_{\text{input}})) \quad (6)$$

其中, F_{AM} 表示轴向注意力路径的输出特征图, F_y 和 F_x 分别表示在垂直轴和水平轴的注意力操作, F_{input} 表示输入特征图 F .

3.5 损失函数

由于非小细胞肺癌的肿瘤体积较小,分割难度较大,不仅需要网络有能够获得包含完整目标结构的能力,且需要更加精细的区域和边界分割,因此,本文采用式(7)定义的混合损失函数(hybrid loss).

$$L_{\text{All}} = L_{\text{BCE}} + L_{\text{IoU}} + L_{\text{SSIM}} \quad (7)$$

L_{All} 表示混合函数, L_{BCE} , L_{IoU} 和 L_{SSIM} 分别表示二元交叉熵 BCE(Binary Cross Entropy) 损失函数^[32], IoU (Intersection over Union) 损失函数^[33] 和结构相似性度量 SSIM (Structural Similarity Index Measurement) 损失

函数^[34].

BCE 损失函数可以很好评估预测值和真实值之间的差距,在二分类和语义分割等领域使用广泛,BCE 损失函数的定义如式(8)所示.

$$L_{\text{BCE}} = - \sum_{(i,j)} \mathbf{G}(i,j) \log(\mathbf{P}(i,j)) - \sum_{(i,j)} (1 - \mathbf{G}(i,j)) \log(1 - \mathbf{G}(i,j)) \quad (8)$$

其中, $\mathbf{G}(i,j) \in \{0,1\}$ 表示像素点 (i,j) 的真实标签, $\mathbf{P}(i,j) \in \{0,1\}$ 表示像素点 (i,j) 的预测值.

IoU 损失函数最初被用作评价两个集合的相似性,同时也可以作为目标检测和语义分割的评价指标,除此之外,还可以用作损失函数,其定义如式(9)所示.

$$L_{\text{IoU}} = 1 - \frac{\sum_{i=1}^H \sum_{j=1}^W \mathbf{G}(i,j) \mathbf{P}(i,j)}{\sum_{i=1}^H \sum_{j=1}^W [\mathbf{P}(i,j) + \mathbf{G}(i,j) - \mathbf{P}(i,j) \mathbf{G}(i,j)]} \quad (9)$$

其中, H 和 W 分别表示图像的高和宽, $\mathbf{G}(i,j) \in \{0,1\}$ 表示像素点 (i,j) 的真实标签, $\mathbf{P}(i,j) \in \{0,1\}$ 表示像素点 (i,j) 的预测值.

结构相似性度量 SSIM 损失函数用来评估图像质量的,它可以有效捕获图像中的结构信息,所以当用作损失函数时可以学习图像中的结构信息. 设 $x = \{x_g: g = 1, \dots, N^2\}$, $y = \{y_p: p = 1, \dots, N^2\}$ 分别是真实标签和预测图中裁剪的两个大小为 $N \times N$ 的对应区域,则 x 和 y 的 SSIM 损失定义为式(10).

$$L_{\text{SSIM}} = 1 - \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (10)$$

其中, $\mu_x, \mu_y, \sigma_x, \sigma_y$ 分别是 x 和 y 的平均值和标准差, σ_{xy} 代表 x 和 y 的协方差,为了避免分母为零,令 $C_1 = 0.01^2, C_2 = 0.03^2$.

在混合损失中,BCE 损失主要是像素级,有助于图像中所有像素的收敛,SSIM 损失是区域级,有助于模型更好的发现结构信息,而 IoU 损失是图像级的,有助于整张图像的特征学习. 本文通过将三个不同的损失函数结合起来,使模型能够获得更完整的目标结构,并获得更好的分割效果.

4 实验结果与分析

4.1 实验环境与评价指标

本文采用开源深度学习框架 PyTorch1.8.0 实现提出的 SOSNet 模型,使用 GPU 加速网络模型的训练和测试,显卡型号为 GeForce RTX 2080Ti. 实验采用 Adam 优化器,初始学习率为 0.001,权重衰减系数为 1×10^{-8} ,

网络初始权重采用 ImageNet 预训练的 ResNet 网络的相应权重. 训练和测试阶段的 batch size 均为 8,训练 1 000 轮次.

实验结果使用 mDice (mean Dice)、mIoU (mean Intersection over Union)、Sensitivity、F1 (F1-Score)、Specificity 和平均绝对误差 MAE (Mean Absolute Error) 这六种常用的医学图像分割评价指标来评价,各评价指标的计算方式如式(11)~(16)所示.

$$\text{mDice} = \frac{1}{k+1} \sum_{i=0}^k \frac{2\text{TP}}{\text{FP} + 2\text{TP} + \text{FN}} \quad (11)$$

$$\text{mIoU} = \frac{1}{k+1} \sum_{i=0}^k \frac{2\text{TP}}{\text{FP} + \text{TP} + \text{FN}} \quad (12)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

$$\text{F1} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (14)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (15)$$

$$\text{MAE} = \frac{1}{k} \sum_{i=1}^k \sum_{j=1}^{N_i} |\hat{y}_j - y_j| \quad (16)$$

在式(11)~(16)中, k 表示分割的类别数,TP (True Positive) 表示网络输出的目标区域是真实的目标区域, TN (True Negative) 表示网络输出的背景区域是真实的背景区域,FP (False Positive) 表示网络输出的目标区域不是真实的目标区域,即错误地将背景分割为目标区域, FN (False Negative) 表示网络输出的背景区域不是真实的背景区域,即错误地将目标区域分割为背景. mDice、mIoU、Sensitivity、F1 和 Specificity 评价指标的取值越大,意味着模型的分割效果越好. 式(16)中, N_i 表示第 i 类的像素点数, \hat{y}_j 表示模型对第 j 个像素点的预测值, y_j 表示第 j 个像素点的真实值, MAE 的值代表模型的平均绝对误差,所以该值越小代表模型越好.

4.2 实验数据集及数据预处理

实验在两个公开数据集上进行,其中,数据集 NSCLC 来自 2020 年第八届 CCF 大数据与计算智能大赛 (CCF Big Data & Computing Intelligence Contest, 简称 CCF BDCI) 的非小细胞肺癌肿瘤自动分割竞赛^[35], 包含 300 例真实的非小细胞肺癌 3D CT 图像数据,包括 CT 图像原始数据和肿瘤标注数据. 我们通过将该数据集中的每个 3D CT 原始数据和肿瘤标注数据进行切片得到每个病例的 2D 数据,经过上述处理后共得到 38 341 张二维 CT 切片;数据集 MSD_L (Medical Segmentation Decathlon-Lung cancer data) 是来自医学图像处理分割十项全能比赛^[36]的肺部肿瘤数据集,包括 96 例非小细胞肺癌患者的术前薄层 3D CT 扫描,经过与 NSCLC 数据集相同的切片处理后,共得到 18 899 张二维 CT 切片. 本文实验按照病例数 7:3 划分训练集和

测试集,并随机抽取训练集样本数量的 10% 作为验证集.

由于非小细胞肺癌肿瘤体积较小,在得到的二维切片中,有很大一部分切片中没有肿瘤存在.因此,本文首先探究不存在分割目标的切片对模型分割结果的影响,具体数据集划分如表 1 所示.其中,*₁表示所有数据转换的切片组成的数据集,*₂表示去除*₁中不含目标的切片后得到的数据集,*₃表示含有目标的切片(正样本)和不含目标的切片(负样本)按照 1:3 组成的数据集.

表 1 实验用数据集详细信息

数据集	病例数	NSCLC			病例数	MSD _L		
		NSCLC ₁	NSCLC ₂	NSCLC ₃		MSD _{L1}	MSD _{L2}	MSD _{L3}
训练集	210	25336	4424	17696	68	12422	1347	5388
测试集	90	10471	814	5235	28	5235	310	1240
验证集	21	2534	442	1770	7	1242	135	540

4.3 SOSNet 非对称编码器-解码器结构性能验证

为了探究本文 SOSNet 主干网络设计的合理性,本小节测试对称和非对称结构对于模型分割效果的影响.使用对称的 BASNet 模型和三种非对称的分割模型在 NSCLC₃ 数据集进行实验测试,结果如表 3 所示.其中,非对称编码器-解码器结构分别是对 BASNet 模型的编码器和解码器进行修改,使之成为不对称的编码器-解码器结构.解码器分别采用 2 种方式修改:5 个双线性上采样逐步还原特征图(即 SOSNet 的解码器)和 10 倍上采样非逐步还原特征图.编码器的改进方式是:采用 2×2 最大池化逐步进行下采样获得特征图,即本文 SOSNet 模型的编码器.编码器和解码器不同时修改.

表 3 对称编解码和不对称编码器-解码器结构对比实验

结构		mDice	mIoU	Sensitivity	F1	Specificity	MAE	
BASNet 主干网		0.688	0.701	0.844	0.813	0.819	0.013	
ED_A	SOSNet 的编码器	0.671	0.703	0.836	0.822	0.815	0.015	
	解码器修改	SOSNet 解码器	0.711	0.732	0.849	0.834	0.830	0.006
		10 倍上采样非逐步还原特征	0.682	0.709	0.846	0.829	0.822	0.007

表 3 结果显示,当编码器不变,修改解码器,采用 5 个双线性上采样逐步还原特征图,即 SOSNet 的解码器,得到非对称的编码器-解码器结构,此时的非对称模型 ED_A 的分割效果最好.其次是原始的编码器-解码器结构完全对称的 BASNet 结构.排名第三的是编码器采用 BASNet 的编码器不变,修改解码器,采用 10 倍上采样非逐步还原特征图,得到不对称的编码器-解码器结构.最差的是编码器采用 2×2 最大池化,逐步进行下采样提取特征图,即 SOSNet 的编码器,解码器是 BASNet 解码器的不对称编码器-解码器结构.

上述结果说明,在编码器-解码器结构中,进行特征提取的编码器部分对于模型分割效果的影响较大,只

为了探究不同数据集对于模型性能的影响,使用 Baseline 模型 BASNet^[29]在表 1 所示的六种不同的数据集进行测试,测试结果如表 2 所示,加粗表示最好结果.

表 2 结果显示,在数据集 NSCLC₃ 与 MSD_{L3} 的实验结果最好,即正负样本比为 1:3 时,模型表现最好,而当使用所有切片组成的数据时,模型表现最差,大量无分割目标的图像对模型来说相当于噪声,说明过多的噪声导致模型效果不佳.因此,本文后续实验将使用 NSCLC₃ 和 MSD_{L3} 数据集来验证模型性能.

表 3 的 BASNet 是编码器-解码器结构对称的模型,ED_A 表示编码器-解码器结构非对称的网络结构,表中加粗数据表示最好结果.

表 2 各个数据集在 Baseline 模型 BASNet 的实验结果

数据集	mDice	mIoU	Sensitivity	F1	Specificity	MAE	
NSCLC	NSCLC ₁	0.601	0.589	0.762	0.811	0.790	0.038
	NSCLC ₂	0.623	0.618	0.784	0.840	0.812	0.003
	NSCLC ₃	0.688	0.701	0.844	0.813	0.819	0.002
MSD	MSD _{L1}	0.639	0.595	0.719	0.749	0.773	0.003
	MSD _{L2}	0.684	0.678	0.857	0.738	0.758	0.005
	MSD _{L3}	0.704	0.698	0.865	0.788	0.778	0.002

有充分提取图像中的特征信息才能更好的进行分割.其次,对于解码器而言,不一定需要完全和编码器对称,只需要能够对特征图进行还原即可,且逐步还原特征图能够有效避免在特征图恢复过程中的信息丢失,使模型达到更好的分割效果.

4.4 ARA 注意力模块性能测试

由于 ARA 模块由两个不同的分支组成,每个分支对于模块的性能都有所影响.为了测试各分支对 ARA 模块性能的影响,将 ARA 模块中的两分支,轴向注意力模块 A 和取反路径模块 R 分别加入 BASNet 中进行消融实验,以验证 ARA 模块每个分支模块的性能.表 4 给出了 ARA 两个分支模块的消融实验结果,实验在 MSD_{L3}

数据集进行,表中加粗表示最好实验结果.

表 4 ARA 模块的两分支模块的消融实验

模型	mDice	mIoU	Sensitivity	F1	Specificity
BASNet+A	0.673	0.698	0.799	0.829	0.817
BASNet+R	0.688	0.725	0.830	0.831	0.823
BASNet+ARA	0.696	0.737	0.837	0.838	0.828

表 4 实验结果显示,分别融合轴向注意力模块 A 和图像取反模块 R 的 BASNet 模型的各项指标均没有融合同时包含二者的 ARA 模块的 BASNet 模型的相应指标好;并且融合图像取反模块 R 的 BASNet 模型的各项指标优于融合轴向注意力模块 A 的 BASNet 模型的性能.

表 4 的实验结果说明,对于小目标的分割,消除背景中的干扰因素对于提升分割效果更为有效,而轴向注意力模块增加了模型的特征提取能力,对于增加分割精度也有一定的作用,综合两者对于提升模型的小目标分割能力具有更好的效果.

ARA 模块的轴向取反去除了非小细胞肺癌肺部 CT 图像中的血管、气管等影响分割效果的组织,只留下分割目标,以提升分割效果.为了验证本文使用的 ARA 模块对小目标肺结节分割的有效性,将 ARA 模块与

ECA (Efficient Channel Attention)^[37]、CBAM (Convolutional Block Attention Module)^[38]、SE (Squeeze-and-Excitation)^[39] 和 SCSE (Spatial and Channel Squeeze & Excitation)^[40] 四种注意力模块分别与 BASNet 基模型结合,进行实验对比,以验证 ARA 模块的优越性能,实验在 MSD₁_3 数据集上进行,实验结果如表 5 所示,加粗表示最好结果.

表 5 结果显示,对非小细胞肺癌 CT 图像的小体积肿瘤结节分割效果,除了 Sensitivity 指标在 BASNet 模型最高外,其他 mDice、mIoU、F1、Specificity 和 MAE 指标均在采用了轴向取反注意力模块 ARA 的 BASNet 模型取得最优值.融合 ECA、CBAM、SE、SCSE 等注意力模块的 BASNet 模型,只在 mIoU 指标上优于基模型 BASNet,在 mDice 和 Sensitivity 指标上均不如基模型 BASNet,F1 和 MAE 指标在融合 CBAM 和 SCSE 注意力模块的 BASNet 模型上优于基模型 BASNet,Specificity 指标仅在融合 SCSE 注意力模块的 BASNet 指标上优于基模型 BASNet.

综合上述分析说明,提出的 ARA 注意力模块能够有效提升 BASNet 模型对非小细胞肺癌 CT 图像的小目标分割效果.

表 5 ARA 模块与其他注意力模块的对比实验结果

模型	mDice	mIoU	Sensitivity	F1	Specificity	MAE
BASNet	0.688	0.701	0.844	0.813	0.819	0.013
BASNet+ECA ^[37]	0.638	0.722	0.823	0.796	0.814	0.017
BASNet+CBAM ^[38]	0.659	0.726	0.823	0.821	0.809	0.012
BASNet+SE ^[39]	0.616	0.715	0.813	0.776	0.807	0.019
BASNet+SCSE ^[40]	0.683	0.732	0.834	0.837	0.820	0.009
BASNet+ARA	0.696	0.737	0.837	0.838	0.828	0.007

4.5 消融实验

为了验证提出的 SOSNet 模型主要组件的有效性,本小节使用 NSCLC_3 数据集通过消融实验对 SOSNet 模型的主要组件进行性能测试.首先采用对称的编码-解码结构网络 BASNet 测试提出的非对称编码-解码器网络 SOSNet 的主要组件 ARA 和 SR,将轴向取反模块 ARA 和结构细化模块 SR 分别加入 BASNet 模型的主干网络,接着将 ARA 和 SR 两模块同时加入 BASNet 模型的主干网络进行实验测试.其次,采用提出的非对称模型 SOSNet 的主干网,即编码器-解码器非对称的 ED_A 模型,分别或者同时加入 ARA 和 SR,测试 SOSNet 主要组件的有效性.测试结果如表 6 所示,其中加粗表示最好结果.

表 6 关于编码器-解码器对称的 BASNet 模型主干网的消融实验结果显示,当 BASNet 的主干网加入 ARA 模块时,其分割性能除了在 Sensitivity 稍有降低外,其他 mDice、mIoU、F1、Specificity 和 MAE 评价指标均得到提升.当对基模型 BASNet 加入结构细化模块

SR 后,网络除了在 Sensitivity 和 Specificity 指标略有下降外,在其他指标 mDice、mIoU、F1 和 MAE 上均得到提升.同时加入 ARA 和 SR 模块后,模型的性能同单独加入 SR 模块类似,除了在 Sensitivity 和 Specificity 指标不如 BASNet 基模型,模型的其他指标 mDice、mIoU、F1 和 MAE 均得到大幅提升.另外,同时加入 ARA 和 SR 模块和单独加入 ARA 或 SR 模块的模型相比,其性能也是在除了 Sensitivity 和 Specificity 指标的其他指标 mDice、mIoU、F1 和 MAE 上均得到提升.

表 6 关于非对称基模型 SOSNet 主干网的消融实验结果显示,除了 Sensitivity 指标,SOSNet 非对称主干网络无论加入轴向取反模块 ARA 还是结构细化模块 SR,模型的分割性能均得到提升,表现在 mDice、mIoU、F1、Specificity 和 MAE 评价指标均优于基模型的相应指标值.另外,加入结构细化模块 SR 的非对称模型的性能在各项指标上均优于加入轴向取反模块 ARA 的非对称模型的性能.当 ARA 模块和 SR 模块同时加入 SOSNet 主干网络时,模型

表 6 不同模块对模型性能影响的消融实验

BaseLine	组件		评价指标					
	ARA	SR	mDice	mIoU	Sensitivity	F1	Specificity	MAE
BASNet 主干网			0.688	0.701	0.844	0.813	0.819	0.013
	√		0.696	0.737	0.837	0.838	0.828	0.007
		√	0.701	0.724	0.843	0.824	0.817	0.007
	√	√	0.717	0.739	0.827	0.848	0.805	0.006
SOSNet 主干网			0.711	0.732	0.849	0.824	0.821	0.006
	√		0.719	0.740	0.832	0.829	0.823	0.006
		√	0.723	0.744	0.844	0.854	0.830	0.005
	√	√	0.737	0.748	0.853	0.865	0.849	0.004

的性能达到最好,各个指标均达到最优,此时的模型即为提出的SOSNet网络.由此可见,提出SOSNet模型非常适宜于用来分割非小细胞肺癌CT图像的小结节.

综上所述可见,对于非小细胞肺癌CT图像的小目标分割,提出的非对称编码-解码结构网络SOSNet具有非常好的分割效果.因此,对非小细胞肺癌CT图像的小结节分割,需要使目标结构更加精细才能获得良好的分割结果.

4.6 SOSNet各组件复杂度分析

作为一种能够端到端实现分割的网络模型,不仅

要准确分割病灶区域,还要有较小的复杂度,能够更快地实现对目标对象的分割任务.因此,通过计算模型的参数量(Parameter)和每秒帧率(Frame Per Second, FPS)来评价提出的非对称编码器-解码器模型SOSNet的各个组件的复杂度.模型的参数量可以衡量模型占用空间的大小,体现了模型的空间复杂度,而每秒帧率表示模型每秒内分割处理的图片数量,可以用来衡量模型的时间消耗,体现了模型的时间复杂度.表7展示了提出的SOSNet模型各组件复杂度分析.

表 7 SOSNet各组件复杂度分析

模型	参数量/MB	每秒帧率/(frame/s)	mDice
BASNet 主干网络	83.03	58	0.688
SOSNet 主干网	40.17	73	0.711
SOSNet 主干网+ARA	81.51	62	0.719
SOSNet 主干网+SR	57.80	66	0.723
SOSNet	90.08	47	0.737

表7中展示了SOSNet模型的各个组件的参数量和每秒帧率以及对NSCLC_3数据集的分割结果的mDice指标值,加粗表示最好结果.表7实验结果显示,本文提出的非对称编解码结构ED_A(即SOSNet主干网)不仅降低了原始模型BASNet模型的参数量,还提升了其分割效率和分割准确率,而将ARA和SR两种不同的组件分别加入非对称编解码结构时,虽然参数量和每秒帧率都没有ED_A网络的好,但准确率均得到了提升.本文提出的SOSNet模型参数量和每秒帧率的表现最差,但分割准确率最优.这说明提出的SOSNet模型用一定的空间和时间复杂度换取了模型的准确率.对医学图像分割而言,准确率对于辅助医生进行疾病诊断具有非常重要的意义,因此,本文提出的SOSNet模型更具有实用价值.

4.7 SOSNet与其他方法的对比实验

为了验证SOSNet模型在非小细胞肺癌分割的有效性,本文将提出的SOSNet模型与其他11种模型进行了比较,这11种模型分别是:U-Net^[3]、U-Net++^[41]、Atten-

tion U-Net^[42]、FCN(Fully Convolutional Networks)^[2]、FusionNet^[43]、SegNet^[7]、CaraNet(Context axial reverse attention Network)^[44]、BASNet^[29]、PraNet(Parallel reverse attention Network)^[45]、COVIDSeg^[46]、PANet(Prior attention Network)^[47].其中,后两种模型是最新的用于肺部CT图像分割的模型.各模型分割结果的mDice、mIoU、Sensitivity、F1、Specificity和MAE各指标的实验结果如表8所示,加粗表示最好结果.另外,我们还用参数量(Parameter)和每秒帧率(FPS)比较了各模型的复杂度,比较结果如表9所示,加粗数据表示最好结果.

表8实验结果显示,提出的SOSNet模型在两个非小细胞肺癌CT图像数据集均取得不错的分割效果,多数情况下优于对比模型.详细分析如下.

在NSCLC_3数据集,分割指标mDice最差的模型是FCN,然后依次分别是SegNet、FusionNet、U-Net、Attention U-Net、UNet++、COVIDSeg、BASNet、CaraNet、PraNet、PANet和SOSNet模型,相对于前七种对比模型,后四种对比模型的分割效果较好,其中BASNet、

表 8 不同方法在两个数据集上的分割实验结果比较

数据集	方法	mDice	mIoU	Sensitivity	F1	Specificity	MAE
NSCLC_3	U-Net ^[3]	0.480	0.407	0.658	0.505	0.635	0.047
	UNet++ ^[41]	0.535	0.479	0.683	0.677	0.706	0.044
	Attention U-Net ^[42]	0.516	0.555	0.657	0.694	0.732	0.029
	FCN ^[2]	0.379	0.368	0.536	0.475	0.554	0.073
	FusionNet ^[43]	0.460	0.459	0.661	0.654	0.709	0.051
	SegNet ^[7]	0.419	0.410	0.610	0.681	0.685	0.057
	CaraNet ^[44]	0.691	0.718	0.782	0.851	0.866	0.012
	BASNet ^[29]	0.688	0.701	0.844	0.813	0.819	0.013
	PraNet ^[45]	0.702	0.695	0.839	0.804	0.826	0.008
	COVIDSeg ^[46]	0.568	0.606	0.727	0.795	0.808	0.022
	PANet ^[47]	0.720	0.717	0.842	0.834	0.825	0.003
	SOSNet	0.737	0.748	0.853	0.865	0.849	0.004
MSD _L -3	U-Net ^[3]	0.329	0.315	0.609	0.665	0.697	0.061
	UNet++ ^[41]	0.498	0.455	0.659	0.682	0.699	0.042
	Attention U-Net ^[42]	0.560	0.497	0.640	0.702	0.727	0.026
	FCN ^[2]	0.314	0.307	0.615	0.626	0.656	0.103
	FusionNet ^[43]	0.438	0.404	0.647	0.658	0.784	0.058
	SegNet ^[7]	0.455	0.424	0.636	0.675	0.766	0.052
	CaraNet ^[44]	0.693	0.685	0.819	0.829	0.773	0.009
	BASNet ^[29]	0.704	0.698	0.865	0.788	0.768	0.025
	PraNet ^[45]	0.697	0.686	0.873	0.782	0.771	0.012
	COVIDSeg ^[46]	0.684	0.665	0.802	0.821	0.765	0.009
	PANet ^[47]	0.721	0.716	0.891	0.831	0.788	0.002
	SOSNet	0.734	0.731	0.888	0.841	0.790	0.004

CaraNet、PraNet 三种模型的各项指标值虽然均不如提出的 SOSNet 模型好,但也均取得了不错的分割效果, mDice 值均在 0.7 左右; PANet 模型的 mDice、mIoU、Sensitivity、F1、Specificity 指标虽然均不如提出的 SOSNet 模型,但其 MAE 指标优于提出的 SOSNet 模型,且是所有模型中最优的. 提出的 SOSNet 模型在指标 mDice、mIoU、Sensitivity、F1 和 Specificity, 比基准模型 BASNet 分别高出 4.9、4.7、0.9、5.2、3.0 个百分点,在平均绝对误差指标 MAE 上,比基准模型 BASNet 提升 0.009.

在 MSD_L-3 数据集,分割指标 mDice 最差的模型也是 FCN, 然后依次分别是 U-Net、FusionNet、SegNet、UNet++、Attention U-Net、COVIDSeg、CaraNet、PraNet、BASNet、PANet 和 SOSNet 模型,其中后五种对比模型的分割性能明显优于比前六种对比模型的分割性能,其中 COVIDSeg、CaraNet 和 PraNet 模型的 mDice 指标接近 0.7, BASNet 和 PANet 模型的 mDice 指标高于 0.7,特别是 PANet 模型的 mDice 指标达到 0.721, 仅比提出的 SOSNet 模型的 mDice 指标低 1.3 个百分点,但 PANet 模型的 Sensitivity 和 MAE 指标优于提出的 SOSNet 模型,且是所有 12 个模型中最优的.

综上各模型分割结果分析可见,提出的非对称小

目标分割模型 SOSNet 能够有效分割出非小细胞肺癌 CT 图像中的小体积肿瘤结节.

表 9 不同方法的复杂度比较

模型	参数量/MB	每秒帧率 /(frame/s)
U-Net	16.47	21
UNet++	34.93	18
Attention U-Net	33.26	14
FCN	19.17	34
FusionNet	77.88	17
SegNet	28.08	20
CaraNet	44.48	24
BASNet	83.03	58
COVIDSeg	12.66	46
PANet	23.14	33
PraNet	31.04	19
SOSNet	90.08	47

表 9 的实验结果可以看出,提出的 SOSNet 模型的参数量是所有模型中最大的,每秒帧率排名第 2,仅次于其基模型 BASNet,但优于其他所有对比模型. 对比表 8 的结果分析,我们可以说本文提出的 SOSNet 模型用模型的

复杂度特别是空间复杂度换取了分割的准确率,即提高了分割准确率,但增加了模型参数,时间效率虽然比不上基模型,不是最优,但也优于其他所有对比模型. 对于医学图像分割而言,准确率是所有指标中最重要.

因此,综合来看,本文提出的针对非小细胞肺癌CT图像分割的非对称编码器-解码器模型 SOSNet 具有很好的实用价值,能有效实现非小细胞肺癌肿瘤结节的分割.

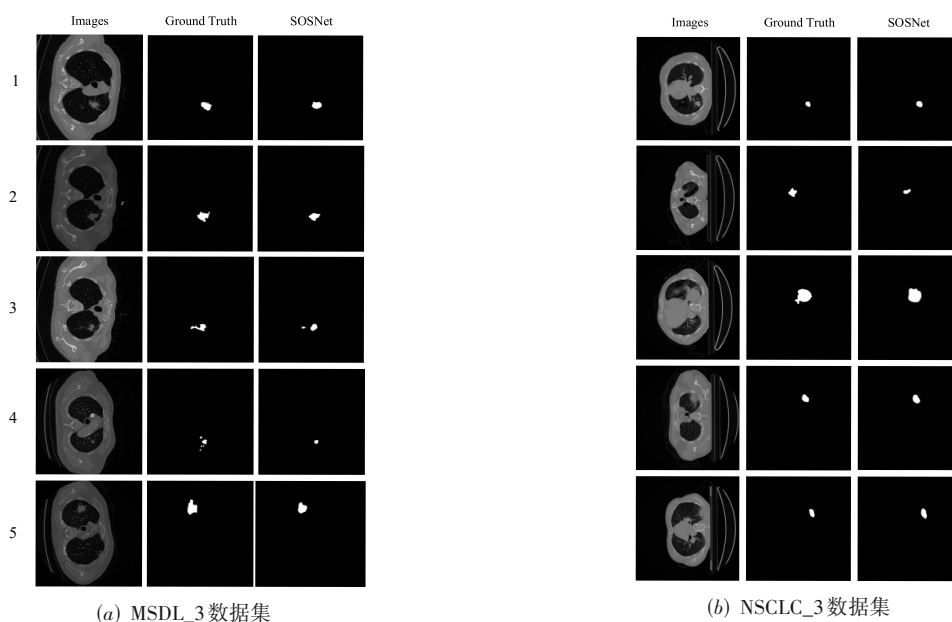


图6 本文 SOSNet 模型对 MSDI_3 和 NSCLC_3 数据集部分样本的分割结果可视化

图6可视化结果显示,提出的非对称分割模型 SOSNet 在两个数据集上的分割结果与图像的真实标记大致相同,对小目标肿瘤分割的整体结果较好,但对一些体积过小且粘连在一起的小结节,如图6(a)MSDI_3数据集的第三和第四个CT图像的小结节,本文 SOSNet 模型只能分割出其中的部分结节,无法分割出体积特别小的结节. 然而,对于独立的小体积结节,提出的 SOSNet 模型能对其进行有效分割,如图6(b)的 NSCLC_3 数据集中第一、第五行的两个CT图像的小体积结节. 另外,对于一些有复杂边界结构的结节,如图6(b)所示的 NSCLC_3 数据集的第3个CT图像,本文 SOSNet 模型的分割结果仍有值得改进的空间.

4.8 结果可视化

为了更加清晰地展示提出的 SOSNet 模型的分割效果,随机在两个数据集中各选择5张图像并对各个模型的分割结果进行可视化,结果如图6所示. 其中图6(a)是 MSDI_3 数据集图像分割结果的可视化,图6(b)是 NSCLC_3 中图像分割结果的可视化. 在两个子图中,第一列是数据集中随机选择的CT图像原图,第二列是真实的分割标签,第三列是提出的 SOSNet 模型的分割结果.

4.9 SOSNet 与 CaraNet 对小目标的分割效率比较

目前针对医学图像小目标分割领域的研究中, CaraNet 模型的性能最好,为了能够更好地展示本文提出的 SOSNet 模型对小目标的分割效率,本小节将特别比较提出的 SOSNet 和 CaraNet 模型^[44]对小目标的分割能力. 表10展示了 SOSNet 和 CaraNet 两种模型对非小细胞肺癌CT图像的小结节目标的分割性能比较,加粗表示最优结果. 图7是该两模型的部分分割结果的可视化展示.

表10的各评价指标结果显示,提出的 SOSNet 模型除了增加了参数量外,其余各项指标,包括分割效果指标和模型效率指标,均优于当前的小目标分割的 SOTA (State Of The Art) 模型 CaraNet,说明提出的 SOSNet 模型是迄今为止最好的医学图像小目标分割模型.

表10 SOSNet 模型与 CaraNet 模型的小目标分割性能比较

方法	mDice	mIoU	Sensitivity	F1	Specificity	MAE	参数量/MB	每秒帧率/(frame/s)
CaraNet	0.693	0.685	0.819	0.829	0.773	0.009	44.48	24
SOSNet	0.734	0.731	0.888	0.841	0.790	0.004	90.08	47

图 7 的可视化结果显示,提出的 SOSNet 模型对小目标的分割结果比 CaraNet 模型更加精确,特别是第一行显示的该两模型对图像 1 的分割结果, CaraNet 模型分割出不是目标的结构,而 SOSNet 模型的分割结果与真实标记基本相同. 此外,提出的 SOSNet 模型比 CaraNet 模型对小目标边界的分割更加精细,更接近于真实标记的目标边界,而 CaraNet 模型的分割结果中,目标区域的边界均是光滑的,与图像中标记的真实目标边界相比,分割结果明显在边界部分不够精细准确.

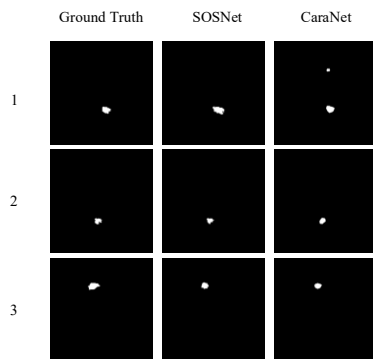


图 7 SOSNet 模型与 CaraNet 模型分割结果可视化

5 结论

提出了一种非对称编码器-解码器结构的非小细胞肺癌肺部 CT 图像的小体积肿瘤结节分割模型 SOSNet, 利用 ResNet 基础层和空洞卷积构造非对称编码器-解码器结构, 即 SOSNet 分割模型的主干网络, 利用轴向取反注意力模块 ARA 逐步擦除背景中对分割有影响的结构, 使用结构细化模块 SR 对主干网络输出的粗略特征图进行结构细化, 从而实现非小细胞肺癌肿瘤结节的分割. 公开数据集的大量实验验证表明, 提出的 SOSNet 模型能有效分割出非小细胞肺癌患者肺部 CT 图像的小目标癌变结节.

虽然 SOSNet 模型可以有效分割出非小细胞肺癌肺部 CT 图像的小体积肿瘤结节, 但对体积过小且相互粘连的非常小的肿瘤小结节和边界结构复杂的肿瘤小结节, 提出的 SOSNet 模型的分割性能还有待进一步提升. 因此, 未来还需在网络结构设计、数据预处理等方面进一步改进, 以便精细化小目标分割效果, 帮助医生实现非小细胞肺癌的及早发现和治疗, 挽救患者生命.

参考文献

- [1] SIEGEL R L, et al. Cancer statistics[J]. CA: A Cancer Journal for Clinicians, 2021, 71(1): 7-33.
- [2] SHELLHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [3] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.
- [4] 谢娟英, 张凯云. XR-MSF-Unet: 新冠肺炎肺部 CT 图像自动分割模型[J]. 计算机科学与探索, 2022, 16(8): 1850-1864.
XIE J Y, ZHANG K Y. XR-MSF-unet: Automatic segmentation model for COVID-19 lung CT images[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(8): 1850-1864. (in Chinese)
- [5] TONG G F, LI Y, CHEN H R, et al. Improved U-NET network for pulmonary nodules segmentation[J]. Optik, 2018, 174: 460-469.
- [6] ROCHA J, CUNHA A, MENDONÇA A M. Conventional filtering versus U-net based models for pulmonary nodule segmentation in CT images[J]. Journal of Medical Systems, 2020, 44(4): 81.
- [7] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [8] SINGADKAR G, MAHAJAN A, THAKUR M, et al. Deep deconvolutional residual network based automatic lung nodule segmentation[J]. Journal of Digital Imaging, 2020, 33(3): 678-684.
- [9] WANG Q, SHEN F Y, SHEN L Y, et al. Lung nodule detection in CT images using a raw patch-based convolutional neural network[J]. Journal of Digital Imaging, 2019, 32(6): 971-979.
- [10] SKOURT B AIT, HASSANI A EL, MAJDA A. Lung CT image segmentation using deep neural networks[J]. Procedia Computer Science, 2018, 127: 109-113.
- [11] BLANC D, RACINE V, KHALIL A, et al. Artificial intelligence solution to classify pulmonary nodules on CT[J]. Diagnostic and Interventional Imaging, 2020, 101(12): 803-810.
- [12] 李亚超, 熊德意, 张民. 神经机器翻译综述[J]. 计算机学报, 2018, 41(12): 2734-2755.
LI Y C, XIONG D Y, ZHANG M. A survey of neural machine translation[J]. Chinese Journal of Computers, 2018, 41(12): 2734-2755. (in Chinese)

- [13] HOLSCHNEIDER M, KRONLAND-MARTINET R, MORLET J, et al. A real-time algorithm for signal analysis with the help of the wavelet transform[C]//Wavelets. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989: 286-297.
- [14] PAPANDREOU G, KOKKINOS I, SAVALLE P A. Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 390-399.
- [15] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[C]//2016 International Conference on Learning Representations(ICLR). Piscataway: IEEE, 2016:1-13.
- [16] CHEN L C, PAPANDREOU G, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[C]//2015 International Conference on Learning Representations (ICLR). Piscataway: IEEE, 2015: 1-14 .
- [17] CHEN L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [18] CHEN L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-12-05) [2023-02-25]. <https://arxiv.org/pdf/1706.05587.pdf>.
- [19] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Computer Vision—ECCV 2018. Cham: Springer International Publishing, 2018: 833-851.
- [20] DAI J F, LI Y, HE K M, et al. R-FCN: Object detection via region-based fully convolutional networks[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: ACM, 2016: 379-387.
- [21] ALEXANDER C, FU C, et al. SSD: Single shot multiBox detector[C]//2016 European Conference on Computer Vision (ECCV). Florence: Springer, 2016: 21-37.
- [22] CHEN K, WANG J, Chen L C, et al. ABC-CNN: An attention based convolutional neural network for visual question answering[EB/OL]. (2016-04-03) [2023-02-25], <https://arxiv.org/pdf/1511.05960.pdf>.
- [23] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//2017 International Conference on Neural Information Processing Systems (NIPS). Vancouver: Curran Associates, Inc., 2017: 6000-6010.
- [24] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7794-7803.
- [25] ZHANG H, GOODFELLOW I, Metaxas D, et al. Self-attention generative adversarial networks[EB/OL]. (2019-06-14)[2023-02-25]. <https://arxiv.org/pdf/1805.08318.pdf>.
- [26] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. New York: ACM, 2014: 2672-2680.
- [27] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 3141-3149.
- [28] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.
- [29] QIN, X, FAN D, et al. Boundary-aware segmentation network for mobile and web applications[EB/OL]. (2021-05-11)[2023-02-25]. <https://arxiv.org/pdf/2101.04704v1.pdf>.
- [30] CHEN S H, TAN X L, WANG B, et al. Reverse attention for salient object detection[C]//Computer Vision—ECCV 2018. Cham: Springer International Publishing, 2018: 236-252.
- [31] WANG H Y, ZHU Y K, GREEN B, et al. Axial-DeepLab: Stand-alone axial-attention for panoptic segmentation[M]//Computer Vision—ECCV 2020. Cham: Springer International Publishing, 2020: 108-126.
- [32] DE BOER P T, KROESE D P, MANNOR S, et al. A tutorial on the cross-entropy method[J]. Annals of Operations Research, 2005, 134(1): 19-67.
- [33] MATTYUS G, LUO W J, URTASUN R. DeepRoadMapper: Extracting road topology from aerial images[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 3458-3466.
- [34] WANG Z, SIMONCELLI E P, BOVIK A C, et al. Multi-

- scale structural similarity for image quality assessment [C]//The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers. Piscataway: IEEE, 2003: 1398-1402.
- [35] 许可乐. 非小细胞肺肿瘤自动分割[EB/OL]. (2020-12-28) [2023-02-25]. <https://www.datafountain.cn/competitions/489>.
- [36] ANTONELLI M, Reinke A, Bakas S, et al. The medical segmentation decathlon[EB/OL]. (2021-06-10) [2023-02-25]. <https://arxiv.org/pdf/2106.05735.pdf>.
- [37] WANG Q L, WU B G, ZHU P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 11531-11539.
- [38] WOO S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//2018 European Conference on Computer Vision (ECCV). Florence: Springer, 2018: 3-19.
- [39] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [40] ROY A G, NAVAB N, WACHINGER C. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks[C]//Medical Image Computing and Computer Assisted Intervention—MICCAI 2018. Cham: Springer International Publishing, 2018: 421-429.
- [41] ZHOU Z, SIDDIQUEE M M R, TAJBAKSH N, et al. UNet++: A nested U-net architecture for medical image segmentation[J]. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2018), 2018, 11045: 3-11.
- [42] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention u-net: learning where to look for the pancreas[EB/OL]. (2018-05-20) [2023-02-25]. <https://arxiv.org/abs/1804.03999.pdf>.
- [43] QUAN T M, HILDEBRAND D G C, JEONG W K. FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics[EB/OL]. (2016-12-16) [2023-02-25]. <https://arxiv.org/ftp/arxiv/papers/1612/1612.05360.pdf>.
- [44] LOU A G, GUAN S Y, Ko H, et al. Caranet: Context axial reverse attention network for segmentation of small medical objects[EB/OL]. (2022-01-13) [2023-02-25]. <https://arxiv.org/abs/2108.07368v1.pdf>.
- [45] FAN D P, JI G P, ZHOU T, et al. PraNet: Parallel reverse attention network for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference. New York: ACM, 2020: 263-273.
- [46] 谢娟英, 夏琴. COVIDSeg: 新冠肺炎肺部 CT 图像轻量化分割模型[J]. 陕西师范大学学报(自然科学版), 2022, 50(3): 65-78.
XIE J Y, XIA Q. COVIDSeg: The lightweight segmentation model for the lung CT images of COVID-19 patients [J]. Journal of Shaanxi Normal University (Natural Science Edition), 2022, 50(3): 65-78. (in Chinese)
- [47] ZHAO X Y, ZHANG P, SONG F, et al. Prior attention network for multi-lesion segmentation in medical images [J]. IEEE Transactions on Medical Imaging, 2022, 41 (12): 3812-3823.

作者简介



谢娟英 女, 1971年4月生于陕西省西安市. 现为陕西师范大学计算机科学学院教授、博士生导师. 获陕西省自然科学二等奖,《中国科学:信息科学》热点论文奖,中国科技期刊卓越行动计划优秀审稿人,入选领跑者F5000、ESI前1%高被引论文等. 主要研究方向为机器学习、数据挖掘、生物医学数据分析等. 在国内外发表学术论文80余篇.

E-mail: xiejuany@snnu.edu.cn



张凯云 女, 1995年10月生于宁夏回族自治区固原市. 陕西师范大学计算机科学学院硕士研究生. 主要研究方向为机器学习、生物医学数据分析.

E-mail: kaiy.zhang@qq.com